Bayes rule
Probabilistic Methods

Reasoning under uncertainly

Bayes theorem

Bayes' Rule (also called Bayes' Theorem) describes how to update the probability of a hypothesis given new evidence. It connects prior beliefs with observed data to produce a posterior probability.

$$P(A \mid B) = (P(B \mid A) \cdot P(A))/ P(B)$$

Bayes theorem: **P(A | B) = (P(B | A) · P(A)) / P(B)**

P(A | B): the *posterior*. The probability of A being true, given that B has occurred. This is what you're solving for.

P(B | A): the *likelihood*. The probability of observing B if A is true.

P(A): the *prior*. Your initial belief about the probability of A, before seeing the evidence.

P(B): the *marginal likelihood* (or evidence). The total probability of observing B under all possible scenarios. It acts as a normalizing constant.

Example: Say a medical test is 99% accurate, a disease affects 1 in 1,000 people, and you test positive. What's the probability you actually have it?

Example: **P(A | B) = (P(B | A) · P(A)) / P(B)**

Example: Say a medical test is 99% accurate, a disease affects 1 in 1,000 people, and you test positive. What's the probability you actually have it?

P(A | B): the *posterior*. The probability of A being true, given that B has occurred. This is what you're solving for.

P(B | A): the *likelihood*. The probability of observing B if A is true.

P(A): the *prior*. Your initial belief about the probability of A, before seeing the evidence.

P(B): the *marginal likelihood* (or evidence). The total probability of observing B under all possible scenarios. It acts as a normalizing constant.

In other words, what is the probability of having the disease if the test is positive?

- P(Disease) = 1/1000
- P(Positive | Disease) =
- P(Positive | No Disease) =
- P(Positive)=
- P(Disease | Positive) =

Example: **P(A | B) = (P(B | A) · P(A)) / P(B)**

Example: Say a medical test is 99% accurate, a disease affects 1 in 1,000 people, and you test positive. What's the probability vou actuallv have it?

P(A | B): the *posterior*. The probability of A being true, given that B has occurred. This is what you're solving for.

P(B | A): the *likelihood*. The probability of observing B if A is true.

P(A): the *prior*. Your initial belief about the probability of A, before seeing the evidence.

P(B): the *marginal likelihood* (or evidence). The total probability of observing B under all possible scenarios. It acts as a normalizing constant.

In other words, what is the probability of having the disease if the test is positive?

- P(Disease) = 1/1000 = 0.001
- P(Positive | Disease) = 99% = 0.99
- P(Positive | No Disease) =
- P(Positive) =
- P(Disease | Positive) =

P(Positive | Disease) + P(Positive | No Disease) =

Example: **P(A | B) = (P(B | A) · P(A)) / P(B)**

P(Positive): What is the overall chance that any random person tests positive?

1. Sick people who correctly test positive
2. Healthy people who mistakenly test positive

Scenario 1: The person is sick AND tests positive

- The chance of being sick is 0.001, because the disease affects 1 in 1,000 people
- If sick, the chance of testing positive is 0.99, because the test is 99%accurate
- the chance of *both* being true is 0.001 × 0.99 = 0.00099 = P(Sick)P(Positive|Sick)

- **Scenario 2: The person is healthy AND tests positive**
  - The chance of being healthy is 0.999, because the disease affects 1 in 1000 people, then (999 in 1,000 are healthy)
  - If healthy, the chance of testing positive is 0.01, because the test is 99% accurate then (1% error)
  - the chance of *both* being true is 0.999 × 0.01 = 0.00999 =P(Healthy)P(Positive|Healthy)

The denominator will be = P(Sick)P(Positive|Sick) + P(Healthy)P(Positive|Healthy) = 0.00099 + 0.00999 = 0.01089

Example: **P(A | B) = (P(B | A) · P(A)) / P(B)**

Example: Say a medical test is 99% accurate, a disease affects 1 in 1,000 people, and you test positive. What's the probability you actually have it?

P(A | B): the *posterior*. The probability of A being true, given that B has occurred. This is what you're solving for.

P(B | A): the *likelihood*. The probability of observing B if A is true.

P(A): the *prior*. Your initial belief about the probability of A, before seeing the evidence.

P(B): the *marginal likelihood* (or evidence). The total probability of observing B under all possible scenarios. It acts as a normalizing constant.

P(Disease) = 1/1000= 0.001; This is the prior. We said the disease affects 1 in 1,000 people. So before any test, the chance a random person has it is 1/1000 = 0.001, or 0.1%.

P(Positive | Disease) = 0.99; This is the likelihood. We said the test is 99% accurate. So if someone *actually has* the disease, the test correctly says "positive" 99% of the time.

P(Positive | No Disease) = 0.01; This is the false positive rate. Since the test is 99% accurate, it gets it wrong 1% of the time. So if someone is *healthy*, there's a 1% chance the test mistakenly says "positive."

P(Positive) = (0.99 × 0.001) + (0.01 × 0.999) = 0.01098; This is the total probability of getting a positive result from *anyone*, sick or healthy. It has two parts:

- 0.99 × 0.001 = 0.00099: This covers the true positives. Out of the 0.1% who are sick, 99% test positive.
- 0.01 × 0.999 = 0.00999: This covers the false positives. Out of the 99.9% who are healthy, 1% mistakenly test positive.

Example: **P(A | B) = (P(B | A) · P(A)) / P(B)**

Example: Say a medical test is 99% accurate, a disease affects 1 in 1,000 people, and you test positive. What's the probability you actually have it?

P(A | B): the *posterior*. The probability of A being true, given that B has occurred. This is what you're solving for.

P(B | A): the *likelihood*. The probability of observing B if A is true.

P(A): the *prior*. Your initial belief about the probability of A, before seeing the evidence.

P(B): the *marginal likelihood* (or evidence). The total probability of observing B under all possible scenarios. It acts as a normalizing constant.

P(Disease) = 1/1000= 0.001; This is the prior. We said the disease affects 1 in 1,000 people. So before any test, the chance a random person has it is 1/1000 = 0.001, or 0.1%.

P(Positive | Disease) = 0.99; This is the likelihood. We said the test is 99% accurate. So if someone *actually has* the disease, the test correctly says "positive" 99% of the time.

P(Positive | No Disease) = 0.01; This is the false positive rate. Since the test is 99% accurate, it gets it wrong 1% of the time. So if someone is *healthy*, there's a 1% chance the test mistakenly says "positive.“

P(Positive) = (0.99 × 0.001) + (0.01 × 0.999) = 0.01098; This is the total probability of getting a positive result from *anyone*, sick or healthy. It has two parts:

- 0.99 × 0.001 = 0.00099: This covers the true positives. Out of the 0.1% who are sick, 99% test positive.
- 0.01 × 0.999 = 0.00999: This covers the false positives. Out of the 99.9% who are healthy, 1% mistakenly test positive.

(Disease | Positive) = (0.99 × 0.001) / 0.01098 ≈ 9%

Bayes Rule: P(X|Y) = (P(Y|X) * P(X)) / P(Y)

Posterior = (Likelihood * Prior) / Evidence

The posterior probability is the probability of the parameters θ given the evidence X: $p(\theta|X)$.

It contrasts with the likelihood function, which is the probability of the evidence given the parameters: $p(X|\theta)$.

Let us have a prior belief that the probability distribution function is $p(\theta)$ and observations x with the likelihood $p(x|\theta)$ probability is defined as

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}.$$ [1]

Example

Suppose there is a mixed school having 60% boys and 40% girls as students. The girls wear trousers or skirts in equal numbers; the boys all wear trousers. An observer sees a (random) student from a distance; all the observer can see is that this student is wearing trousers. What is the probability this student is a girl? The correct answer can be computed using Bayes' theorem.

The event G is that the student observed is a girl, and the event T is that the student observed is wearing trousers. To compute the posterior probability $P(G|T)$, we first need to know:

Suppose there is a mixed school having 60% boys and 40% girls as students. The girls wear trousers or skirts in equal numbers; the boys all wear trousers. An observer sees a (random) student from a distance; all the observer can see is that this student is wearing trousers. What is the probability this student is a girl? The correct answer can be computed using Bayes' theorem.

The event G is that the student observed is a girl, and the event T is that the student observed is wearing trousers. To compute the posterior probability $P(G|T)$, we first need to know:

- $P(G)$, or the probability that the student is a girl regardless of any other information. Since the observer sees a random student, meaning that all students have the same probability of being observed, and the percentage of girls among the students is 40%, this probability equals 0.4.
- $P(B)$, or the probability that the student is not a girl (i.e. a boy) regardless of any other information (B is the complementary event to G). This is 60%, or 0.6.
- $P(T|G)$, or the probability of the student wearing trousers given that the student is a girl. As they are as likely to wear skirts as trousers, this is 0.5.
- $P(T|B)$, or the probability of the student wearing trousers given that the student is a boy. This is given as 1.
- $P(T)$, or the probability of a (randomly selected) student wearing trousers regardless of any other information. Since $P(T) = P(T|G)P(G) + P(T|B)P(B)$ (via the law of total probability), this is $P(T) = 0.5 \times 0.4 + 1 \times 0.6 = 0.8$.

$$\mathrm{P(G \mid T) = (0.5 \times 0.4) / 0.8}$$

$$\mathrm{P(G \mid T) = 0.2 / 0.8 = \mathbf{0.25 \text{ or } 25\%}}$$

```
P(X|Y) = ( P(Y|X) * P(X) ) / P(Y)
Posterior = ( Likelihood * Prior ) / Evidence
```

- Suppose we have 100 movies and 50 books.

- There are 3 different movie types: Action, Sci-fi, Romance,
- 2 different book types: Sci-fi, Romance

- 20 of those 100 movies are Action.
- 30 are Sci-fi
- 50 are Romance.

- 15 of those 50 books are Sci-fi and 35 are Romance

```
P(X|Y) = ( P(Y|X) * P(X) ) / P(Y)
Posterior = ( Likelihood * Prior ) / Evidence
```

prior/likelihood/posterior/evidence


```
P(X|Y) = ( P(Y|X) * P(X) ) / P(Y)
Posterior = ( Likelihood * Prior ) / Evidence
```


- Suppose we have 100 movies and 50 books.

- There are 3 different movie types: Action, Sci-fi, Romance,
- 2 different book types: Sci-fi, Romance

- 20 of those 100 movies are Action.
- 30 are Sci-fi
- 50 are Romance.

- 15 of those 50 books are Sci-fi and 35 are Romance

- The probability that it's a movie P(Movie) is: ___
- The probability that it's a book is P(Book): ___
- The probability that it's a Sci-fi type is P(Sci-fi): __
- The probability that it's a action type P(action): ____
- The probability that it's a romance type P(romance): ____

- If we already know it's a movie, then the probability that it's an action movie is P(action | movie) =?, P(sci-fi | movie) =?, P(romance | movie) =?

- If we already know it's a book, then that probability that it's an Sci-fi book movie is P(action | book) =?, P(sci-fi | book) =?, P(romance | book) =?
- P(movie | action) = p(action | movie)p(action)/p(movie)

The probability that it's a movie is 100/150, 50/150 for book.

The probability that it's a Sci-fi type is 45/150,

The probability that it's a Action type is 20/150

The probability that it's a Romance type is 85/150

If we already know it's a movie, then

the probability that it's an action movie is 20/100
the probability that it's an Sci-fi movie is 30/100,
the probability that it's an Romance movie is 50/100,

If we already know it's a book, then
the probability that it's an Sci-fi book is 15/50
the probability that it's an Romance book is 35/50

Here, P(movie|Sci-fi) is called **Posterior**,

P(Sci-fi|Movie) is **Likelihood**,

P(movie) is **Prior**,

P(Sci-fi) is **Evidence**.

Prior: **Before** we **observe** it's a Sci-fi type, the object is completely unknown to us. Our goal is to find out the possibility that it's a movie, we actually have the data **prior(or before)** our **observation**, which is the possibility that it's a movie if it's a completely unknown object: **P(movie)**.

Posterior: **After** we **observed** it's a Sci-fi type, we know something about the object. Because it's **post(or after)** the **observation**, we call it **posterior**: P(movie|Sci-fi).

Evidence: Because we've already known it's a Sci-fi type, what has happened is happened. We **witness** it's appearance, so to us, it's an **evidence**, and the chance we get this evidence is **P(Sci-fi)**.

Likelihood: The dictionary meaning of this word is chance or probability that one thing will happen. Here it means when it's a movie, what the chance will be if it is also a Sci-fi type. This term is very important in Machine Learning.

Naïve Bayes

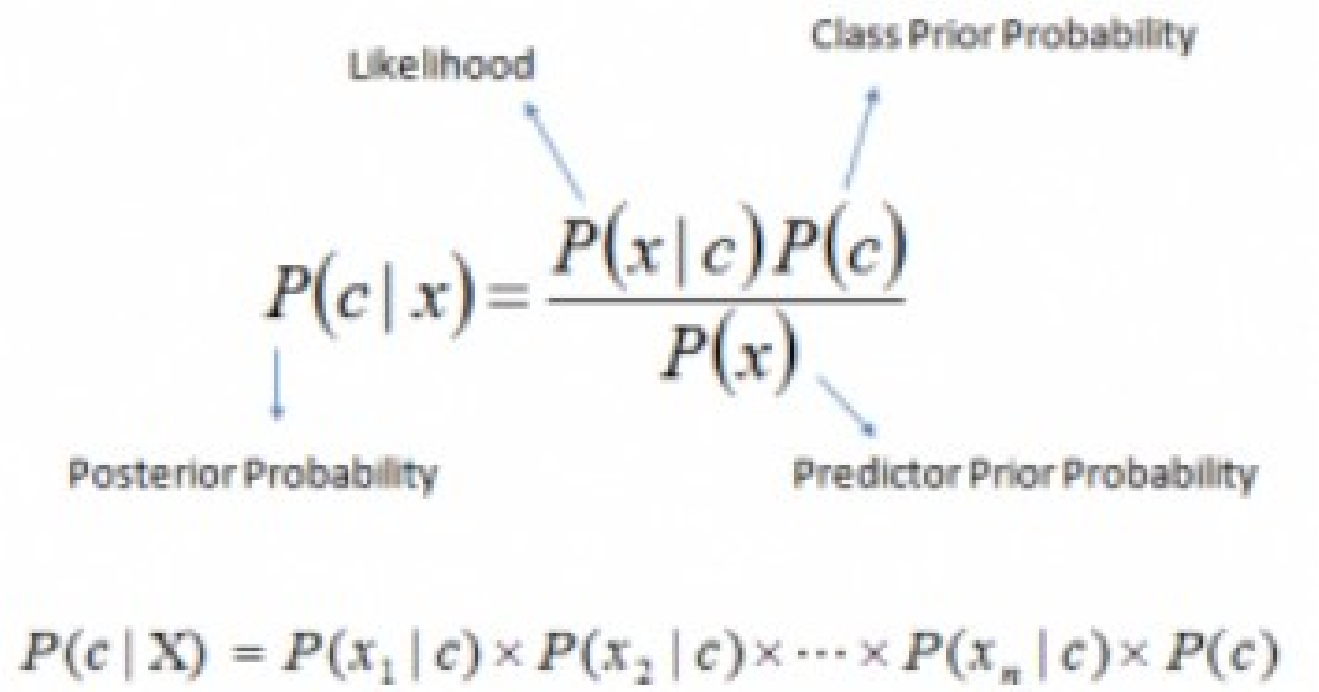
Strong (naive) independence assumptions between the features.

Naïve Base

- based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

- For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple, and that is why it is known as 'Naive'

Naïve Bayes supervised learning



$$P(c \mid \mathbf{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Example- Given the following dataset, train NB

$$P(c \mid \mathbf{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

$$P(C_k \mid x_1, x_2, \ldots, x_n) = \frac{P(C_k) \times \prod_{i=1}^{n} P(x_i \mid C_k)}{P(x_1, x_2, \ldots, x_n)}$$

# day samples	Weather	Wind	Humidity	Label (Play tennis?)
1	Sunny	Weak	High	No
2	Sunny	Strong	High	No
3	Cloudy	Weak	High	Yes
4	Rainy	Weak	High	Yes
5	Rainy	Weak	Normal	Yes
6	Rainy	Strong	Normal	No
7	Cloudy	Strong	Normal	Yes
8	Sunny	Weak	High	No
9	Sunny	Weak	Normal	Yes
10	Rainy	Weak	Normal	Yes

Given a new observation (Weather=Sunny, Wind=Weak, Humidity=Normal), predict if this person will play tennis

P(play tennis = yes/no | Weather=Sunny, Wind=Weak, Humidity=Normal) ?

Example- Given the following dataset, tr

$$P(C_k \mid x_1, x_2, \ldots, x_n) = \frac{P(C_k) \times \prod_{i=1}^{n} P(x_i \mid C_k)}{P(x_1, x_2, \ldots, x_n)}$$

Step 1 — Prior probabilities P(Class)

- Yes = 6 rows → P(Yes) = 6/10 = **0.60**
- No = 4 rows → P(No) = 4/10 = **0.40**

Step 2 — Likelihood tables P(feature | Class)

For each feature, count how often each value appears within each class:

Weather	P(·|Yes)	P(·|No)
Sunny	1/6 = 0.17	3/4 = 0.75
Cloudy	2/6 = 0.33	0/4 = 0.00
Rainy	3/6 = 0.50	1/4 = 0.25

Wind	P(·|Yes)	P(·|No)
Weak	5/6 = 0.83	2/4 = 0.50
Strong	1/6 = 0.17	2/4 = 0.50

Humidity	P(·|Yes)	P(·|No)
High	2/6 = 0.33	3/4 = 0.75
Normal	4/6 = 0.67	1/4 = 0.25

# day samples	Weather	Wind	Humidity	Label (Play tennis?)
1	Sunny	Weak	High	No
2	Sunny	Strong	High	No
3	Cloudy	Weak	High	Yes
4	Rainy	Weak	High	Yes
5	Rainy	Weak	Normal	Yes
6	Rainy	Strong	Normal	No
7	Cloudy	Strong	Normal	Yes
8	Sunny	Weak	High	No
9	Sunny	Weak	Normal	Yes
10	Rainy	Weak	Normal	Yes

Given a new observation (Weather=Sunny, Wind=Weak, Humidity=Normal), predict if this person will play tennis

Example- Given the following dataset, t

$$P(C_k \mid x_1, x_2, \ldots, x_n) = \frac{P(C_k) \times \prod_{i=1}^{n} P(x_i \mid C_k)}{P(x_1, x_2, \ldots, x_n)}$$

Given a new observation (Weather=Sunny, Wind=Weak, Humidity=Normal), predict if this person will play tennis

- P(Yes | Weather=Sunny, Wind=Weak, Humidity=Normal)= P(Yes) × P(Sunny|Yes) × P(Weak|Yes) × P(Normal|Yes) = 0.60 × 0.17 × 0.83 × 0.67 = 0.0564
- P(No | Weather=Sunny, Wind=Weak, Humidity=Normal)= P(No) × P(Sunny|No) × P(Weak|No) × P(Normal|No) = 0.40 × 0.75 × 0.50 × 0.25 = 0.0375

Denominator (the normalizer): [Score(Yes) + Score(No)] = 0.0564 + 0.0375 = 0.0939

P(Yes | features) = Score(Yes) / [Score(Yes) + Score(No)] = 0.0564/0.0939 = 0.60

P(No | features) = Score(No) / [Score(Yes) + Score(No)] = 0.0375/0.0939 = 0.39

Notice they **add up to exactly 1.0** , that's the entire job of the denominator.

It rescales the raw scores so they become valid probabilities

Weather	P(·|Yes)	P(·|No)
Sunny	1/6 = 0.17	3/4 = 0.75
Cloudy	2/6 = 0.33	0/4 = 0.00
Rainy	3/6 = 0.50	1/4 = 0.25
Wind	**P(·|Yes)**	**P(·|No)**
Weak	5/6 = 0.83	2/4 = 0.50
Strong	1/6 = 0.17	2/4 = 0.50
Humidity	**P(·|Yes)**	**P(·|No)**
High	2/6 = 0.33	3/4 = 0.75
Normal	4/6 = 0.67	1/4 = 0.25

Exercise

Given a new observation (Weather=Cloudy, Wind=Weak, Humidity=Normal), predict if this person will play tennis

$$P(C_k \mid x_1, x_2, \ldots, x_n) = \frac{P(C_k) \times \prod_{i=1}^{n} P(x_i \mid C_k)}{P(x_1, x_2, \ldots, x_n)}$$

Weather	P(·|Yes)	P(·|No)
Sunny	1/6 = 0.17	3/4 = 0.75
Cloudy	2/6 = 0.33	0/4 = 0.00
Rainy	3/6 = 0.50	1/4 = 0.25
Wind	**P(·|Yes)**	**P(·|No)**
Weak	5/6 = 0.83	2/4 = 0.50
Strong	1/6 = 0.17	2/4 = 0.50
Humidity	**P(·|Yes)**	**P(·|No)**
High	2/6 = 0.33	3/4 = 0.75
Normal	4/6 = 0.67	1/4 = 0.25

Programming use apply Laplace smoothing

For programming purposes, we use the Laplace smoothing to fix the problem of the following example

Given a new observation (Weather=Cloudy, Wind=Weak, Humidity=Normal), predict if this person will play tennis

Weather	P(·|Yes)	P(·|No)
Sunny	1/6 = 0.17	3/4 = 0.75
Cloudy	2/6 = 0.33	0/4 = 0.00
Rainy	3/6 = 0.50	1/4 = 0.25

P(Cloudy | Yes) = 2/6 = 0.333
P(Cloudy | No) = 0/4 = 0.000 ← problem!

Denominator = 0.1109 + 0.0000 = 0.1109
P(Yes | features) = 0.1109 / 0.1109 = 100%
P(No | features) = 0.0000 / 0.1109 = 0%

Fix it using Laplace smoothing

Weather (V=3)	With Laplace (Yes)	With Laplace (No)
Sunny	(1+1)/(6+3)=0.222	(3+1)/(4+3)=0.571
Cloudy	(2+1)/(6+3)=0.333	**(0+1)/(4+3)=0.143**
Rainy	(3+1)/(6+3)=0.444	(1+1)/(4+3)=0.286

Wind (V=2)	P(·|Yes)	P(·|No)
Weak	(5+1)/(6+2)=0.750	(2+1)/(4+2)=0.500
Strong	(1+1)/(6+2)=0.250	(2+1)/(4+2)=0.500

Humidity (V=2)	P(·|Yes)	P(·|No)
High	(2+1)/(6+2)=0.375	(3+1)/(4+2)=0.667
Normal	(4+1)/(6+2)=0.625	(1+1)/(4+2)=0.333

Example

$$P(C_k \mid x_1, x_2, \ldots, x_n) = \frac{P(C_k) \times \prod_{i=1}^{n} P(x_i \mid C_k)}{P(x_1, x_2, \ldots, x_n)}$$

Create a classifier to decide if you will play given a sunny Day.

P(play = yes | weather = sunny) =?

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

P(Yes | Sunny) = (P(Sunny | Yes) * P(Yes))/ P (Sunny)

P(No | Sunny) = P(Sunny | No) * P(No) / P (Sunny)

P(Yes | Sunny) = P(Sunny | Yes) * P(Yes) / P (Sunny)

Here we have

P (Sunny |Yes) = 3/9 = 0.33, P(Sunny) = 5/14 = 0.36, P(Yes)= 9/14 = 0.64

Now,

P (Yes | Sunny) = 0.33 * 0.64 / 0.36 = 0.60, which has higher probability than No

P (Yes | Sunny) > P (No | Sunny) Then P (Yes | Sunny)

2 Car theft Example

Attributes are Color , Type , Origin, and the subject, stolen can be either yes or no.

2.1 data set

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

We want to classify a Red Domestic SUV